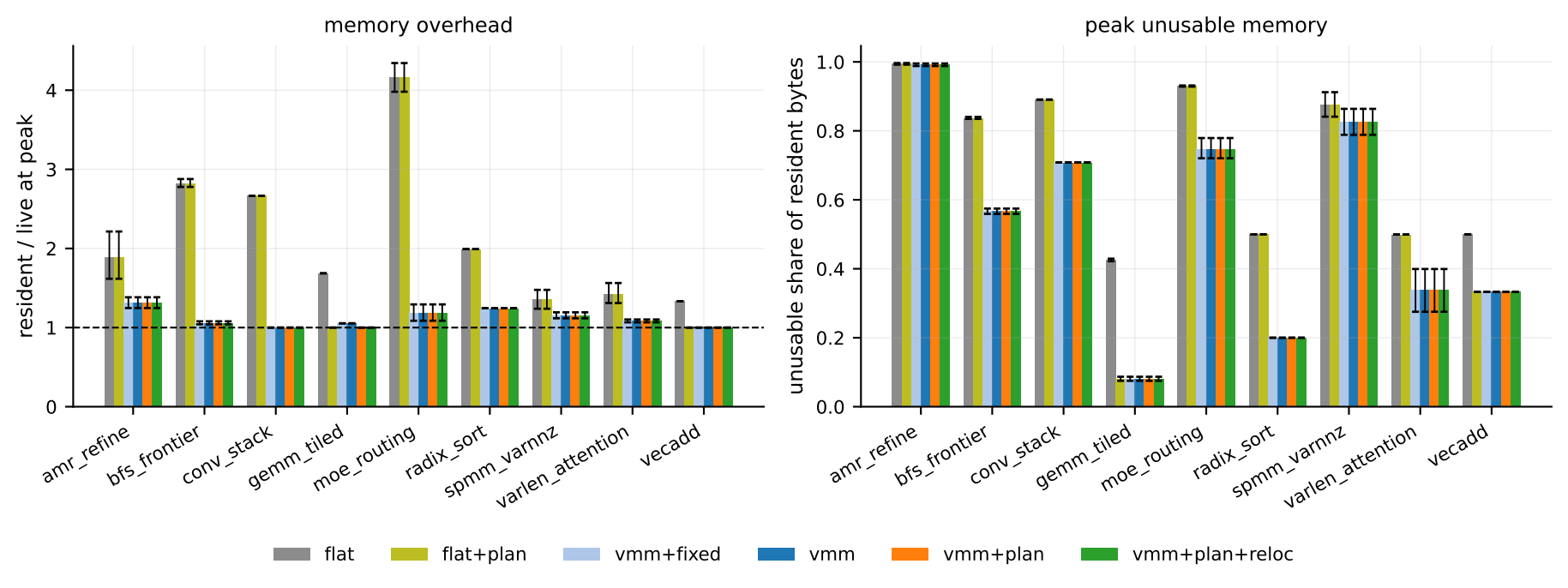
memory overhead
resident / live at peak
0
1
2
3
4
amr_refine
bfs_frontier
conv_stack
gemm_tiled
moe_routing
radix_sort
spmm_varnnz
varlen_attention
vecadd
peak unusable memory
unusable share of resident bytes
0.0
0.2
0.4
0.6
0.8
1.0
amr_refine
bfs_frontier
conv_stack
gemm_tiled
moe_routing
radix_sort
spmm_varnnz
varlen_attention
vecadd
flat
flat+plan
vmm+fixed
vmm
vmm+plan
vmm+plan+reloc